

# Identifying Causal Effects With Proxy Variables of an Unmeasured Confounder

BY WANG MIAO

*Beijing International Center for Mathematical Research, Peking University, 5 Summer Palace Road, Haidian District, Beijing 100871, P.R.C.*

mwfy@pku.edu.cn

ZHI GENG

*School of Mathematical Sciences, Peking University, 5 Summer Palace Road, Haidian District, Beijing 100871, P.R.C.*

zhigeng@pku.edu.cn

AND ERIC TCHETGEN TCHETGEN

*Department of Biostatistics, Harvard University, 677 Huntington Avenue, Boston, Massachusetts 02115, U.S.A.*

etchetge@hsph.harvard.edu

## SUMMARY

Suppose we are interested in a causal effect that is confounded by an unobserved variable. Suppose however one has available proxy variables of the confounder. We show that, with at least two independent proxy variables satisfying a certain rank condition, the causal effect is nonparametrically identified, even if the error mechanism, i.e., the conditional distribution of the proxies given the confounder, may not be identified. Our result generalizes the identification strategy of Kuroki & Pearl (2014) that rests on identification of the error mechanism. When only one proxy of the confounder is available, or the required rank condition is not met, we develop a strategy to test the null hypothesis of no causal effect.

*Some key words:* Confounder; Identification; Measurement error; Proxy.

## 1. INTRODUCTION

Unmeasured confounding is a crucial problem in observational studies. Simpson's paradox (Simpson, 1951) is an elegant illustration of the type of bias that may arise in causal inference subject to unmeasured confounding. Sometimes, an analyst may have access to one or more proxies of the unobserved confounder, for example, a mismeasured version of the confounder. In this case, it may seem natural to directly adjust for the available proxies in order to reduce bias due to unobserved confounding (Greenland, 1980, 1996; Carroll et al., 2006; Ogburn & Vanderweele, 2013; Kuroki & Pearl, 2014). Greenland (1980) suggested that adjustment by a binary nondifferential proxy, which is independent of the treatment and the outcome after conditioning on the confounder, generally reduces bias due to confounding; for a polytomous confounder, certain monotonicity assumptions are indispensable to guarantee such bias attenuation (Ogburn & Vanderweele, 2012; Ogburn & Vanderweele, 2013). But even when the monotonic-

ity assumptions are met, the approach of Greenland (1980) and Ogburn & Vanderweele (2013) is unable to completely eliminate confounding bias. Greenland & Lash (2008) developed a matrix adjustment method that can completely account for unobserved confounding, however, the approach requires external information on the error mechanism, i.e., conditional distribution of the proxy given the confounder, and therefore cannot be applied in routine situations where the error mechanism is unknown. Fortunately, when multiple proxies of the confounder are available, as shown by Kuroki & Pearl (2014), one can sometimes identify the error mechanism and thus the causal effect without external information. Kuroki & Pearl (2014) studied identification of the causal effect with two independent proxies in the context of graph-based models encoding certain conditional independencies about the proxies, where identification means that the causal effect can be determined uniquely from the joint distribution of observed variables. Figure 1 presents several plausible causal diagrams with proxies of the confounder, where  $X$  and  $Y$  denote the treatment and the outcome respectively,  $U$  denotes the confounder that is not observed, but two proxies  $Z$  and  $W$  of  $U$  are available. For graph-theoretic terminology, we refer readers to Pearl (2009). Table 1 presents the corresponding conditional independencies for the diagrams in Figure 1. Diagram (a) corresponds to the nondifferential proxy (Carroll et al., 2006; Greenland & Lash, 2008), and (b)–(c) allow the proxy to depend on the treatment and the outcome respectively; (d) depicts a situation where two proxies are independent conditional on the true confounder, and (e)–(f) allow the proxies to depend on certain observed variables. Practical examples for such diagrams can be found in Carroll et al. (2006), Greenland & Lash (2008) and Kuroki & Pearl (2014). Using the  $\text{do}(x)$  operator of Pearl (2009), the causal effect of  $X$  on  $Y$  is  $\text{pr}\{y \mid \text{do}(x)\} = \sum_u \text{pr}(y \mid x, u)\text{pr}(u)$ , where the symbol  $\text{pr}$  stands for probability. For (d)–(e), Kuroki & Pearl (2014) establish sufficient conditions for nonparametric identification of  $\text{pr}(w \mid u)$ , which suffices to identify  $\text{pr}\{y \mid \text{do}(x)\}$  by applying the matrix adjustment technique of Greenland & Lash (2008).

Model (f) is more general than (d)–(e), but only under a joint normal model, Kuroki & Pearl (2014) established identification of the causal effect for (f). The nonparametric identification method of Kuroki & Pearl (2014) that depends on identification of the error mechanism  $\text{pr}(w \mid u)$ , does not apply to (f), because in general,  $\text{pr}(w \mid u)$  is not identifiable for (f). We illustrate this with a counterexample in section 2. In the current literature, nonparametric identification of the causal effect for (f) is not available, which greatly limits its application. However, we propose a novel strategy to nonparametrically identify the causal effect for (f) without identifying  $\text{pr}(w \mid u)$ . We consider a categorical confounder in section 2, and then we generalize the identification strategy to the continuous case in section 3. The required condition is weaker than that of Kuroki & Pearl (2014). Moreover, when only one proxy is available, i.e., (a)–(c), or the proposed identification condition required for (d)–(f) is not met, we establish that it is nonetheless sometimes possible to obtain a valid empirical test of the null hypothesis of no causal effect. Here, by virtue of focusing on identification, asymptotic analyses are only given consideration, whereby results are strictly given for a hypothetical super-population model.

Table 1: Conditional independencies of causal diagrams

(a) $Z \perp\!\!\!\perp (X, Y) \mid U$ ,	(b) $Z \perp\!\!\!\perp Y \mid (U, X)$ ,
(c) $W \perp\!\!\!\perp X \mid U$ ,	(d) $W \perp\!\!\!\perp (Z, X, Y) \mid U, Z \perp\!\!\!\perp (X, Y) \mid U$ ,
(e) $W \perp\!\!\!\perp (Z, X, Y) \mid U, Z \perp\!\!\!\perp Y \mid (U, X)$ ,	(f) $W \perp\!\!\!\perp (Z, X) \mid U, Z \perp\!\!\!\perp Y \mid (U, X)$ .

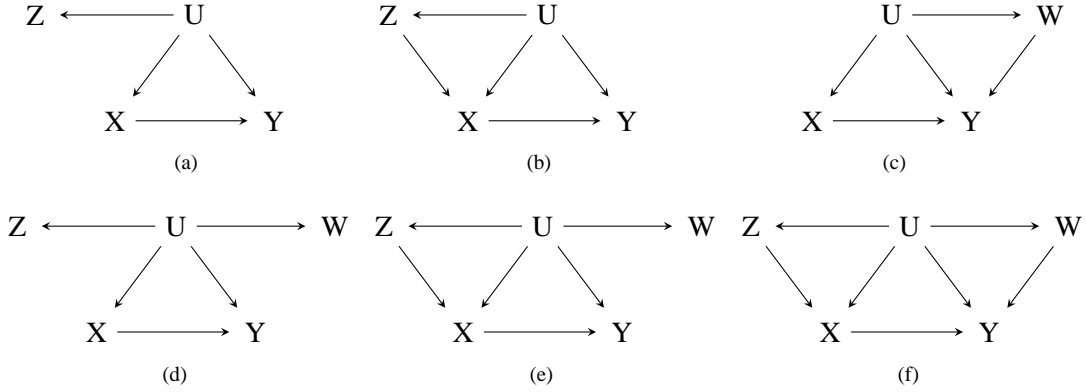


Fig. 1: Causal diagrams with confounder proxies.

## 2. IDENTIFICATION WITH A CATEGORICAL CONFOUNDER

As (d) and (e) can be viewed as special cases of (f) with  $W \perp\!\!\!\perp Y \mid U$ , we focus on identification of the causal effect for (f). Suppose that  $W, Z$  and  $U$  are discrete variables, each with  $k$  categories. For notational convenience, we use  $P(W \mid u) = \{\text{pr}(w_1 \mid u), \dots, \text{pr}(w_k \mid u)\}^T$ ,  $P(w \mid U) = \{\text{pr}(w \mid u_1), \dots, \text{pr}(w \mid u_k)\}$  and  $P(W \mid U) = \{P(W \mid u_1), \dots, P(W \mid u_k)\}$  to denote a column vector, a row vector and a matrix that consist of conditional probabilities of  $W$  given  $U$  respectively. For other variables, vectors and matrices consisting of conditional probabilities are analogously defined:  $P(U \mid z, x) = \{\text{pr}(u_1 \mid z, x), \dots, \text{pr}(u_k \mid z, x)\}^T$ ,  $P(U \mid Z, x) = \{P(U \mid z_1, x), \dots, P(U \mid z_k, x)\}$ ,  $P(y \mid Z, x) = \{\text{pr}(y \mid z_1, x), \dots, \text{pr}(y \mid z_k, x)\}$ ,  $P(y, W \mid z, x) = \{\text{pr}(y, w_1 \mid z, x), \dots, \text{pr}(y, w_k \mid z, x)\}^T$  and  $P(y, W \mid Z, x) = \{P(y, W \mid z_1, x), \dots, P(y, W \mid z_k, x)\}$ . Diagram (f) implies  $W \perp\!\!\!\perp (Z, X) \mid U$  and  $Z \perp\!\!\!\perp Y \mid (U, X)$ , thus, we have

$$P(W \mid Z, x) = P(W \mid U)P(U \mid Z, x), \quad (1)$$

$$P(y \mid Z, x) = P(y \mid U, x)P(U \mid Z, x). \quad (2)$$

Based on (1) and (2), we identify the causal effect under the condition:

(i)  $P(W \mid Z, x)$  is invertible for all  $x$ .

Condition (i) is weaker than that of Kuroki & Pearl (2014), which substantially requires both  $P(W \mid Z, x)$  and  $P(y, W \mid Z, x)$  to be invertible for all  $x$  and  $y$ . Condition (i) can be tested with observed variables; it is in fact equivalent to say both  $P(W \mid U)$  and  $P(U \mid Z, x)$  are invertible (Theorem 1 in the Appendix), which formalizes the idea that both  $W$  and  $Z$  are informative proxies. Under (i), (1)–(2) imply

$$\begin{aligned} P(U \mid Z, x) &= P(W \mid U)^{-1}P(W \mid Z, x), \\ P(y \mid Z, x) &= P(y \mid U, x)P(W \mid U)^{-1}P(W \mid Z, x), \end{aligned} \quad (3)$$

and we have

$$P(y \mid U, x) = P(y \mid Z, x)P(W \mid Z, x)^{-1}P(W \mid U). \quad (4)$$

From (4), identification of  $P(y \mid U, x)$  does depend on the error mechanism  $P(W \mid U)$ , but after taking expectation over  $U$  on both sides of (4), it turns out that identification of  $\text{pr}\{y \mid \text{do}(x)\}$

does not depend on identification of  $P(W | U)$ . To see this, let  $P(U) = \{\text{pr}(u_1), \dots, \text{pr}(u_k)\}^T$  and  $P(W)$  is analogously defined, then  $\text{pr}\{y | \text{do}(x)\} = P(y | U, x)P(U)$  is identified by

$$\text{pr}\{y | \text{do}(x)\} = P(y | Z, x)P(W | Z, x)^{-1}P(W). \quad (5)$$

In contrast, the approach of Kuroki & Pearl (2014) depends on identification of  $P(W | U)$ , which rests on the assumption  $W \perp\!\!\!\perp Y | U$ . When  $W \not\perp\!\!\!\perp Y | U$ , i.e., for model (f), however,  $P(W | U)$  is in general not identified, and therefore, their approach fails even though the causal effect may still be identified by our formula (5). Thus, we have in fact relaxed their more stringent assumption. We illustrate with a binary example.

*Example 1.* Suppose the true data generating mechanism  $\text{pr}(x, y, u, z, w)$  is encoded in the following probability matrices:

$$P(X) = \begin{pmatrix} 5 \\ 5 \end{pmatrix} / 10, \quad P(Z | X) = \begin{pmatrix} 14 & 9 \\ 7 & 12 \end{pmatrix} / 21, \quad P(U | Z, x_i) = \begin{pmatrix} 8 - 2i & 9 - i \\ 2 + 2i & 1 + i \end{pmatrix} / 10,$$

$$P(y_1, W | U, x_i) = \begin{pmatrix} 6i & 22i - 14 \\ 24i + 16 & 4i + 20 \end{pmatrix} / 100, \quad P(y_2, W | U, x_i) = \begin{pmatrix} 20 - 6i & 54 - 22i \\ 64 - 24i & 40 - 4i \end{pmatrix} / 100,$$

for  $i = 1, 2$ . Then we have  $P(W | U, x_i) = P(W | U)$ , thus, the data generating process satisfies diagram (f). The distribution of observed variables is captured by  $P(X)$ ,  $P(Z | X)$  and  $P(y, W | Z, x) = P(y, W | U, x)P(U | Z, x)$  for  $y = y_1, y_2$  and  $x = x_1, x_2$ . Letting

$$A = \begin{pmatrix} 1.1 & -0.4 \\ -0.1 & 1.4 \end{pmatrix},$$

we construct a new data generating process  $\text{pr}_2(x, y, u, w, z)$  encoded in the following matrices:  $P_2(U | Z, x_i) = A^{-1}P(U | Z, x_i)$ ,  $P_2(y_1, W | U, x_i) = P(y_1, W | U, x_i)A$ , and  $P_2(y_2, W | U, x_i) = P(y_2, W | U, x_i)A$  for  $i = 1, 2$ , together with  $P_2(X) = P(X)$  and  $P_2(Z | X) = P(Z | X)$ . The new data generating process satisfies (f) with  $P_2(W | U, x_i) = P_2(W | U) = P(W | U)A$ , thus,  $P_2(W | U) \neq P(W | U)$ . However, the distribution of observed variables remains the same, because for all  $x$  and all  $y$ ,

$$P_2(y, W | Z, x) = P_2(y, W | U, x)P_2(U | Z, x) = P(y, W | Z, x).$$

Thus,  $P(W | U)$  cannot be identified. But applying (5), the causal effect is identified:  $\text{pr}\{y_1 | \text{do}(x_i)\} = 0.3i + 1.04$  for  $i = 1, 2$ , which can be verified from  $\text{pr}(x, y, u, z, w)$ .

For the binary case, the right hand side of (5) reduces to a more explicit form:

$$\frac{\{\text{pr}(w_1) - \text{pr}(w_1 | z_2, x)\}\text{pr}(y | z_1, x)}{\text{pr}(w_1 | z_1, x) - \text{pr}(w_1 | z_2, x)} + \frac{\{\text{pr}(w_1 | z_1, x) - \text{pr}(w_1)\}\text{pr}(y | z_2, x)}{\text{pr}(w_1 | z_1, x) - \text{pr}(w_1 | z_2, x)},$$

which is a weighted average of  $\text{pr}(y | z_i, x)$ ,  $i = 1, 2$ . It can be viewed as a modified version of the adjustment formula  $\sum_{i=1}^2 \text{pr}(y | z_i, x)\text{pr}(z_i)$  previously suggested by Greenland (1980) and Ogburn & Vanderweele (2013). Their approach can incorporate only one proxy, and is biased for  $\text{pr}\{y | \text{do}(x)\}$  due to confounding. As a second proxy  $W$  is available, instead of the weight  $\text{pr}(z)$ , we use a modified weight that can eliminate the bias due to imperfect adjustment by  $Z$ . If  $Z$  is indeed the true confounder, then both adjustment formulas are equivalent and reduce to a common expression for the causal effect.

If  $W$  and  $Z$  have more categories than  $U$ , the causal effect is identifiable as long as  $P(W | Z, x)$  has rank  $k$ . Identification for this case is achieved by using respective coarsening variables

$W'$  and  $Z'$  with  $k$  categories such that  $P(W' | Z', x)$  is invertible. In empirical studies, the unobserved confounder is sometimes continuous, in which case,  $k = +\infty$  and the full rank condition (i) is not valid. We further establish identification of the causal effect for the continuous case.

### 3. IDENTIFICATION WITH A CONTINUOUS CONFOUNDER

Under (f), we generalize (5) to the continuous case by generalizing the rank condition (i) to the following completeness conditions: for all square-integrable function  $g$  and for all  $x$ ,

- (ii)  $E\{g(u) | z, x\} = 0$  almost surely if and only if  $g(u) = 0$  almost surely;
- (iii)  $E\{g(z) | w, x\} = 0$  almost surely if and only if  $g(z) = 0$  almost surely.

Conditions (ii) and (iii) can accommodate both categorical and continuous confounders; they require both  $W$  and  $Z$  to have at least as many categories as  $U$ . For a categorical confounder with categorical proxy variables, (ii) and (iii) must hold under (i); for a continuous one, we suppose  $Z$  and  $W$  are continuous, then many commonly-used models satisfy (ii) and (iii), such as the generalized linear models (Shao, 2003, Proposition 2.1, page 110) and location-scale families (2011 technical report by Hu, Yingyao and Shiu, Ji-Liang). Letting  $f$  denote the density function of a continuous variable, instead of the matrix form (3) for the categorical case, identification for the continuous case involves the solution  $h(w, x, y)$  to the integral equation:

$$\text{pr}(y | z, x) = \int_{-\infty}^{+\infty} h(w, x, y) f(w | z, x) dw \text{ for all } x, y \text{ and } z. \quad (6)$$

Existence of the solution is guaranteed by (iii), and for such a solution  $h(w, x, y)$ , we must have  $\text{pr}(y | u, x) = \int_{-\infty}^{+\infty} f(w | u) h(w, x, y) dw$  under (ii). We provide the details in Theorem 2 of the Appendix. Because  $f(w | z, x)$  and  $\text{pr}(y | z, x)$  are available from observed variables,  $h(w, x, y)$  can be obtained from observed variables by solving (6). The causal effect is identified by

$$\text{pr}\{y | \text{do}(x)\} = \int_{-\infty}^{+\infty} \text{pr}(y | u, x) f(u) du = \int_{-\infty}^{+\infty} f(w) h(w, x, y) dw. \quad (7)$$

We consider a normal model to illustrate, in this case,  $h(w, x, y)$  has an explicit form.

*Example 2.* Assume (f), and suppose that  $f(z, u, w | x)$  follows a joint normal density for all  $x$ ; then  $f(z | w, x)$ ,  $f(w | z, x)$  and  $f(u | z, x)$  must be normal densities, and (ii)–(iii) hold. Suppose one has available  $\text{pr}(y | z, x)$  and  $f(w | z, x) = 1/\sigma(x)\phi\{(w - \beta_0(x) - \beta_1(x)z)/\sigma(x)\}$  from observed variables, with  $\phi$  the density function of a standard normal variable. Then  $h(w, x, y)$  is available in integral form:

$$h(w, x, y) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp\left\{\frac{ivw}{\sigma(x)}\right\} \frac{h_2(v, x, y)}{h_1(v)} dv,$$

where  $i = (-1)^{1/2}$  denotes the imaginary unity,  $h_1(v)$  and  $h_2(v, x, y)$  are Fourier transforms of  $\phi$  and  $\text{pr}(y | z, x)$  respectively:  $h_1(v) = \int_{-\infty}^{+\infty} \exp(-ivz)\phi(z)dz$ , and

$$h_2(v, x, y) = \frac{\beta_1(x)}{\sigma(x)} \int_{-\infty}^{+\infty} \exp\left\{-iv\frac{\beta_0(x) + \beta_1(x)z}{\sigma(x)}\right\} \text{pr}(y | z, x) dz.$$

After obtaining  $h(w, x, y)$ , we can further identify  $\text{pr}\{y | \text{do}(x)\}$  from (7).

Example 2 is a generalization of the joint normal model for  $(X, Y, U, W, Z)$  considered by Kuroki & Pearl (2014), however, here we allow  $X$  and  $Y$  to follow an arbitrary distribution,

not necessarily normal. Under a joint normal model for  $(X, Y, U, W, Z)$ , the result of example 2 can be further strengthened: one may first apply linear regression or maximum likelihood estimation to the observed variables to obtain  $f(y | z, x) \sim N(\alpha_0 + \alpha_1 z + \alpha_2 x, \sigma_1^2)$ ,  $f(w | z, x) \sim N(\beta_0 + \beta_1 z + \beta_2 x, \sigma_2^2)$  and  $f(w)$ ; then one may apply the result of example 2 to obtain  $h(w, x, y)$ ; and finally  $\text{pr}\{y | \text{do}(x)\}$  is a normal density  $N(\gamma_0 + \gamma_1 x, \sigma^2)$  with  $\gamma_1 = \alpha_1 \beta_2 / \beta_1 - \alpha_2$  and  $(\gamma_0, \sigma^2)$  presented in the Supplementary Material. One can also verify that the path coefficient  $\partial E(y | u, x) / \partial x = \gamma_1$ . This result is indeed consistent with that of Kuroki & Pearl (2014), which is obtained by variance analysis of the joint normal model. Such details are relegated to the Supplementary Material.

#### 4. HYPOTHESIS TESTING WITH PROXY VARIABLES

Conditions (i)–(iii) implicitly require that both  $W$  and  $Z$  have more categories than  $U$ , otherwise, the causal effect is in general not identifiable. We provide a counterexample in the Supplementary Material. Nevertheless, as we elaborate below, when one of the proxies has fewer categories than  $U$ , we may use such proxy variables to test the null hypothesis of no causal effect of  $X$  on  $Y$ . We consider a categorical  $U$  with  $k$  levels, and focus on testing

$$\mathbb{H}_0 : X \perp\!\!\!\perp Y \mid U$$

in model (f), then we generalize the results to other diagrams. The null hypothesis means that  $X$  has no causal effect on  $Y$  at any level of  $U$ , which is equivalent to  $\text{pr}(y | u) = \text{pr}(y | u, x)$ , and is also equivalent to  $\text{pr}(x | u) = \text{pr}(x | u, y)$  for all  $x, y$  and  $u$ . Thus, rejection of the null hypothesis is evidence in favor of causation between  $X$  and  $Y$ . Because  $U$  is not observed, we cannot directly check the divergence between  $\text{pr}(y | u, x)$  and  $\text{pr}(y | u)$ . We develop a measure for  $\text{pr}(y | u, x) - \text{pr}(y | u)$  based on proxy variables of  $U$ . We assume

- (iv)  $X, Z$  and  $W$  have  $i, j$  and  $k$  categories respectively, with  $ij \geq k + 1$ ; and the matrix  $Q = \{P(W | Z, x_1), \dots, P(W | Z, x_i)\}$  has full row rank (i.e., the row vectors are linearly independent).

Under (iv),  $P(W | U)$  is invertible (Theorem 1 in the Appendix), and (3) still holds:  $P(y | Z, x) = P(y | U, x)P(W | U)^{-1}P(W | Z, x)$  for all  $x$ . We have the decomposition:

$$\begin{aligned} P(y | Z, x) &= P(y | U)P(W | U)^{-1}P(W | Z, x) \\ &\quad + \{P(y | U, x) - P(y | U)\}P(W | U)^{-1}P(W | Z, x). \end{aligned} \quad (8)$$

As  $x$  varies, (8) reveals two separate sources of variability of  $P(y | Z, x)$ :  $P(W | Z, x)$  and  $P(y | U, x) - P(y | U)$ . Under  $\mathbb{H}_0$ ,  $P(W | Z, x)$  must be the only source of variability of  $P(y | Z, x)$  because  $P(y | U, x) = P(y | U)$ . Thus, we can test  $\mathbb{H}_0$  by checking whether  $P(W | Z, x)$  explains away variability in  $P(y | Z, x)$ . Letting  $q = \{P(y | Z, x_1), \dots, P(y | Z, x_i)\}$ , and

$$\xi = q - qQ^T(QQ^T)^{-1}Q,$$

i.e., the least-square residual by regressing  $q$  against  $Q$ , then  $\xi$  measures the residual variability of  $P(y | Z, x)$  that  $P(W | Z, x)$  does not empirically explain away, i.e., variability of  $P(y | U, x) - P(y | U)$ . Under  $\mathbb{H}_0$ ,  $\xi = (0, \dots, 0)$ , thus, we can check how far away  $\xi$  is from zero to assess whether  $\mathbb{H}_0$  is true.

The testing strategy proposed above is readily generalized to accommodate polytomous  $W$  with more than  $k$  levels by using an appropriate coarsening of  $W$  to construct the above met-

ric. The inequality  $ij \geq k + 1$  is essential; otherwise, variability of  $P(y | Z, x)$  is dominated by  $P(W | Z, x)$ . For  $ij = k$ ,  $\xi$  is identically zero even when  $\mathbb{H}_0$  does not hold.

As a counterpart, we can also test  $\mathbb{H}_0$  with a measure of variability of  $P(x | U, y) - P(x | U)$ . This is achieved by swapping  $W$  and  $Z$ , and swapping  $X$  and  $Y$  whenever they appear in (iv) and in the construction of  $\xi$ .

The proposed test for (f) applies to (d)–(e). For (b)–(c), we can test  $\mathbb{H}_0$  by treating one of the proxies as degenerate, with point mass at an arbitrary level, i.e.,  $W$  is degenerate in model (b) and  $Z$  is degenerate in (c) respectively. In (b)–(c), to guarantee  $ij \geq k + 1$ ,  $Y$  and  $X$  are required to have more categories than  $k$ , respectively. The result equally applies to model (a), which is a special case of (b) and (c). We illustrate with the following hypothetical example.

*Example 3.* Consider (d) with binary  $U, W, Y$ , degenerate  $Z$ , and ternary  $X$ , and suppose the data generating mechanism is determined by the following probability matrices:

$$\begin{aligned} P(X) &= (2, 3, 5)^T / 10, & P(U | X) &= \begin{pmatrix} 4 & 9 & 2 \\ 6 & 1 & 8 \end{pmatrix} / 10, & P(W | U) &= \begin{pmatrix} 8 & 5 \\ 2 & 5 \end{pmatrix} / 10, \\ P(Y | U, x_1) &= \begin{pmatrix} 5 & 4 \\ 5 & 6 \end{pmatrix} / 10, & P(Y | U, x_2) &= \begin{pmatrix} 4 & 6 \\ 6 & 4 \end{pmatrix} / 10, & P(Y | U, x_3) &= \begin{pmatrix} 3 & 8 \\ 7 & 2 \end{pmatrix} / 10. \end{aligned}$$

The causal effect is  $\text{pr}\{y_1 | \text{do}(x_i)\} = (6.5i + 38)/100$  for  $i = 1, 2, 3$ . From the observed variables, we obtain

$$Q = \begin{pmatrix} 62 & 77 & 56 \\ 38 & 23 & 44 \end{pmatrix} / 100, \quad q = (44, 42, 70),$$

and thus  $\xi = (-11, 3, 8)/100$ , which provides evidence to reject  $\mathbb{H}_0$ .

Testing the hypothesis  $X \perp\!\!\!\perp Y | U$  has been a central topic in causal inference, since Simpson's paradox was revealed. The proposed method provides a response to Simpson's paradox. We assess the plausibility of a causal claim subject to confounding by describing necessary conditions for the null hypothesis to hold, and we reject the null hypothesis if such conditions are violated. The approach is along with the spirit of sensitivity analysis initiated by Cornfield et al. (1959) that relates different causative hypotheses to sensitivity parameters, and assesses their plausibility by checking whether extreme values of the sensitivity parameters are plausible. However, our approach directly employs the observed variables instead of hypothetical values for sensitivity parameters that are not observed. Pearl (2010) also suggested an approach to test  $\mathbb{H}_0$  in the presence of an unobserved confounder, which rests on a known error mechanism, i.e., conditional distribution of the proxy given the unobserved confounder. In contrast, our method does not involve such external information. Furthermore, our method only employs certain conditional distributions of the observed variables, but not their joint distribution.

## 5. DISCUSSION

The proposed identification strategy can apply to the treatment effect on the treated  $\text{pr}\{y | \text{do}(x), x'\} = \sum_u \text{pr}(y | u, x) \text{pr}(u | x')$ , by replacing the marginal distribution of  $W$  with the conditional distribution of  $W$  given  $X = x'$  in the identification formulas for  $\text{pr}\{y | \text{do}(x)\}$ . Identifiability of the diagrams reflects the well-known fact in latent class analysis that a latent class model is identified with at least three independent proxies of the latent factor (Kruskal, 1976; Goodman, 1974; Allman et al., 2009). However, our analysis for (f) highlights that certain parameter of interest, such as the causal effect, is still identifiable with only two independent proxies, even though the latent class model is not completely identified. Besides for measurement

error and latent class analysis, this work has potential application in observational studies when negative controls are available (Lipsitch et al., 2010; Sofer et al., 2016), in which case, a negative control outcome that is not causally affected by the treatment, and a negative control exposure that does not causally affect the primary outcome, may serve as proxies of the confounder. The proposed methods can accommodate observed covariates, by simply incorporating them in all conditional densities when targeting a conditional causal effect, and marginalizing over them in the conditional causal effect to obtain a marginal causal effect. Although not discussed in this paper, inference by maximum likelihood is fairly straightforward and follows from the plug-in principle, when the causal effect is identified. More robust semiparametric methods of inference for settings where a large vector  $X$  of covariates must be adjusted for, will be explored elsewhere.

#### ACKNOWLEDGEMENT

The work is partially supported by the China Scholarship Council and U.S.A. National Institute of Health.

#### SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes details for examples 1–2, and a counterexample to identification of (a).

#### APPENDIX

**THEOREM 1.** *Suppose matrices  $A, B$  and  $C$  satisfy  $A = BC$ , then the rank of  $A$  does not exceed the rank of  $B$  or  $C$ .*

It is a fundamental theorem of matrix algebra (Banerjee & Roy, 2014, Theorem 5.4, page 136), and thus, we omit the proof. Applying Theorem 1 to matrices  $P(W | Z, x)$ ,  $P(W | U)$  and  $P(U | Z, x)$  that satisfy  $P(W | Z, x) = P(W | U)P(U | Z, x)$ , the ranks of both  $P(W | U)$  and  $P(U | Z, x)$  are at least as large as that of  $P(W | Z, x)$ .

Identification for the continuous case rests on the following theorem:

**THEOREM 2.** *Suppose  $K(s, t)$  is a bivariate function that satisfies the completeness condition: for all square-integrable function  $g$ ,  $\int_{-\infty}^{+\infty} K(s, t)g(s)ds = 0$  almost surely if and only if  $g = 0$  almost surely; then*

- (A1) *for  $\int_{-\infty}^{+\infty} K(s, t)g(s)ds = \int_{-\infty}^{+\infty} K(s, t)g'(s)ds$ , we must have  $g = g'$ ;*
- (A2) *the range of the integral transform  $R(K) = \{\int_{-\infty}^{+\infty} K(s, t)h(t)ds : h(t) \text{ is square integrable}\}$ , must include all square integrable functions of  $s$ .*

*Proof.* The proof of (A1) is trivial. For  $\int_{-\infty}^{+\infty} K(s, t)g(s)ds = \int_{-\infty}^{+\infty} K(s, t)g'(s)ds$ , we have  $\int_{-\infty}^{+\infty} K(s, t)\{g(s) - g'(s)\}ds = 0$ , and by the completeness condition,  $g - g' = 0$ .

The proof of (A2) rests on the lemma about orthogonal decomposition of a Hilbert space:

**LEMMA 1.** *Let  $H$  be a Hilbert space,  $\Omega$  a closed subspace of  $H$ ,  ${}^\perp\Omega$  the orthogonal complement of  $\Omega$ , then every element of  $H$  can be decomposed uniquely as a sum of a vector in  $\Omega$  and in  ${}^\perp\Omega$ .*

The lemma is a part of Lax (2002, Theorem 3, page 55). According to Lemma 1, in order to prove (A2), we only need to prove that the orthogonal complement of  $R(K)$  is  $\{g = 0\}$ . For any  $g$  belonging to the



orthogonal complement of  $R(K)$ , letting  $g'(t) = \int_{-\infty}^{+\infty} K(v, t)g(v)dv$ , we have

$$\begin{aligned} 0 &= \int_{-\infty}^{+\infty} g(s) \left\{ \int_{-\infty}^{+\infty} K(s, t)g'(t)dt \right\} ds = \int_{-\infty}^{+\infty} g(s) \int_{-\infty}^{+\infty} K(s, t) \left( \int_{-\infty}^{+\infty} K(v, t)g(v)dv \right) dt ds \\ &= \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} K(s, t)g(s)ds \right)^2 dt, \end{aligned}$$

where the first equality follows from the definition of orthogonal complement, the second one is obtained by substitution of  $g'(t)$ , and the third one is obtained by exchanging the integration order with respect to  $s$  and  $t$ . Thus,  $\int_{-\infty}^{+\infty} K(s, t)g(s)ds = 0$ , and by (A1), we have  $g = 0$ . We have completed the proof.  $\square$

We show how this theorem applies to the identification with a continuous confounder. For fixed  $(x, y)$ , according to (A2), the solution  $h(w, x, y)$  to (6) must exist under (iii). For such a solution, (6) implies  $\int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{+\infty} f(w | u)f(u | z, x)du \right\} h(w, x, y)dw = f(y | z, x)$ . Because  $f(y | z, x) = \int_{-\infty}^{+\infty} f(y | u, x)f(u | z, x)du$ , under (ii), we have  $\text{pr}(y | u, x) = \int_{-\infty}^{+\infty} f(w | u)h(w, x, y)dw$  by applying (A1).

#### REFERENCES

- ALLMAN, E. S., MATIAS, C. & RHODES, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics* **37**, 3099–3132.
- BANERJEE, S. & ROY, A. (2014). *Linear Algebra and Matrix Analysis for Statistics*. Boca Raton, FL: Taylor & Francis.
- CARROLL, R., RUPPERT, D., STEFANSKI, L. & CRAINICEANU, C. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. Boca Raton: Chapman & Hall/CRC, 2nd ed.
- CORNFIELD, J., HAENSZEL, W., HAMMOND, E. C., LILIENFELD, A. M., SHIMKIN, M. B. & WYNDER, E. L. (1959). Smoking and lung cancer: recent evidence and a discussion of some questions. *Journal of the National Cancer Institute* **22**, 173–203.
- GOODMAN, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* **61**, 215–231.
- GREENLAND, S. (1980). The effect of misclassification in the presence of covariates. *American Journal of Epidemiology* **112**, 564–569.
- GREENLAND, S. (1996). Basic methods for sensitivity analysis of biases. *International Journal of Epidemiology* **25**, 1107–1116.
- GREENLAND, S. & LASH, T. (2008). Bias analysis. In *Modern Epidemiology*, K. Rothman, S. Greenland & T. Lash, eds. Philadelphia: Lippincott Williams and Wilkins, 3rd ed., pp. 345–380.
- KRUSKAL, J. B. (1976). More factors than subjects, tests and treatments: an indeterminacy theorem for canonical decomposition and individual differences scaling. *Psychometrika* **41**, 281–293.
- KUROKI, M. & PEARL, J. (2014). Measurement bias and effect restoration in causal inference. *Biometrika* **101**, 423–437.
- LAX, P. (2002). *Functional Analysis*. New York: Wiley.
- LIPSITCH, M., TCHETGEN TCHETGEN, E. & COHEN, T. (2010). Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* **21**, 383–388.
- OGBURN, E. L. & VANDERWEELE, T. J. (2012). On the nondifferential misclassification of a binary confounder. *Epidemiology* **23**, 433–439.
- OGBURN, E. L. & VANDERWEELE, T. J. (2013). Bias attenuation results for nondifferentially mismeasured ordinal and coarsened confounders. *Biometrika* **100**, 241–248.
- PEARL, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge, MA: Cambridge University Press., 2nd ed.
- PEARL, J. (2010). On measurement bias in causal inference. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence*. Catalina Island, pp. 425–432.
- SHAO, J. (2003). *Mathematical Statistics*. New York: Springer, 2nd ed.
- SIMPSON, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society. Series B* **13**, 238–241.
- SOFER, T., RICHARDSON, D. B., COLINCINO, E., SCHWARTZ, J. & TCHETGEN TCHETGEN, E. J. (2016). On simple relations between difference-in-differences and negative outcome control of unobserved confounding. *Statistical Science*. To appear.